



ARTICLE

An Unpaired Dual-Domain Image Dehazing Framework Using Unsupervised Learning

Shunpeng Yang¹, Yunpeng Wu¹, Wenwen Qin¹, Cheng Yang^{2,*} and Yu Qian³

¹Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, China

²TravelSky Technology Limited, Beijing, China

³Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC, USA

*Corresponding Author: Cheng Yang. Email: yangcheng@travelsky.com.cn

Received: 18 December 2025; Accepted: 03 February 2026; Published: 18 May 2026

ABSTRACT: To enhance traffic infrastructure health monitoring via computer vision (CV) in adverse weather conditions, image dehazing has emerged as a critical processing step. However, current supervised dehazing models, typically trained on synthetic hazy-clean image pairs, often demonstrate limited generalization ability when deployed in real-world haze scenarios. This study proposes a novel unsupervised dehazing framework named the unpaired dual-domain dehazing network (UD³Net). Initially, a novel dual-domain convolutional mixer (DCM) is developed, which can extract local features in the spatial domain and global features in the frequency domain to achieve thorough information fusion, aiming to facilitate accurate estimation of physical parameters in haze imaging. Then, a dual-domain adaptive gating (DAG) fusion module using an attention mechanism is also designed to dynamically integrate both the spatial domain and frequency domain semantics for image dehazing. Secondly, a newly developed multi-prior contrastive loss (MPC) is proposed to supervise intrinsic properties of unpaired data in space-frequency domains, in order to reduce the loss of semantic information in unpaired unsupervised dehazing. Finally, extensive experiments conducted on both synthetic datasets and real haze datasets validate that the proposed model surpasses the state-of-the-art unsupervised dehazing approaches yet remains while maintaining robust generalization capabilities in real haze scenarios.

KEYWORDS: Image dehazing; unsupervised learning; dual-domain; multi-prior; Fourier transform

1 Introduction

Haze is a harmful weather phenomenon that endangers human health and also disrupts optical imaging. During hazy conditions, light undergoes refraction or scattering due to airborne particles, resulting in image blurring and reduced contrast in camera captures. The image quality degradation caused by haze makes further image processing challenging, particularly for algorithms such as target detection and semantic segmentation that require high image quality. Haze images reduce the accuracy of these algorithms, thereby affecting the application of computer vision in industrial inspections, infrastructure maintenance, and environmental change detection. Consequently, image dehazing is a very urgent and important image processing task [1,2].

As comprehensively summarized in [3], existing methods can be broadly categorized into prior-based physical models and deep learning-based approaches. Early image dehazing methods [4–9] primarily depended on manually crafted priors and conventional atmospheric scattering model theories.

However, their limited generalization capabilities resulted in suboptimal performance in challenging haze conditions. Recently, numerous deep learning-based image dehazing methods [10–14] have demonstrated commendable performance on various image dehazing benchmarks. Nonetheless, those supervised learning approaches require a substantial number of paired images and are restricted to training on synthetic data. The substantial variability among different hazy image domains presents a challenge for supervised dehazing methods, which often exhibit limited generalization capabilities. Most importantly, collecting paired data of large-scale haze and clear images in the real world is extremely impractical [15].

Recently, the image dehazing method using unsupervised learning has received some attention [16–21]. Typical dehazing network [16,22] without paired data, which utilize the CycleGAN framework to facilitate direct mutual conversion between hazy and clear images. However, neglecting the physical constraints inherent to haze imaging can lead to model collapse when trained on unpaired data [23]. Some physically-based unpaired dehazing methods [17,19,24] decompose the transmission map into density and depth. However, estimation errors associated with these physical parameters can significantly compromise the stability and robustness of the dehazing results. Additionally, several studies [18,21,23] have integrated contrastive learning to boost the performance of unpaired dehazing networks; however, relying solely on spatial domain information proves inadequate to impose effective constraints on the dehazing output. Specifically, existing unsupervised dehaze approaches face the following significant challenges:

- (1) The image dehazing task requires the synergistic utilization of local details and global information. Traditional Convolutional Neural Networks (CNNs) are inherently limited to capturing features of local regions, making it difficult to effectively model global contextual correlations. In contrast, although the self-attention mechanism possesses inherent advantages in handling long-range dependencies and global information, it introduces considerable computational overhead [25]. Current mainstream methods lack efficient global information modeling mechanisms, posing significant challenges to striking a balance between performance and efficiency. Furthermore, when simple element-wise addition is adopted for feature fusion in the backbone network, issues such as feature redundancy and receptive field mismatch are prone to arise, which restricts the improvement of model performance.
- (2) In the field of unsupervised image dehazing, incorporating contrastive loss to optimize dehazing performance is an effective technical strategy. However, due to the lack of ground truth in this scenario, most existing methods [18,19,26] randomly select clear images as positive samples and randomly input hazy images as negative samples during training. The model is optimized to make the dehazed output features approach positive samples while moving away from negative samples. This sample construction method tends to cause confusion between dehazing-related features and irrelevant features in the feature space, thereby affecting the visual quality and fidelity of the generated images. Although a few methods [23,25] have re-examined the construction logic of positive and negative samples, they fail to fully leverage image priors, ultimately limiting the improvement of their dehazing performance.

This study provides a novel unpaired dual-domain dehaze framework (UD³Net) to address the aforementioned issues, which is built on the atmospheric scattering model [27] and has better interpretability than learning-based dehazing methods. To address the first challenge mentioned above, this paper proposes a Dual-Domain Convolutional Mixer (DCM), which is constructed based on the Fast Fourier Transform (FFT). The DCM aims to balance local and global feature modeling capabilities while ensuring efficient model inference performance. Meanwhile, a Dual-Domain Adaptive Gating Fusion Module (DAG) is designed to achieve the complementary fusion of spatial and frequency domain features. Inspired by the relevant studies [13,28,29], this research leverages abundant image priors to tackle the second challenge: at the

frequency-domain level, the input hazy image is decoupled into haze-related features (amplitude component) and haze-irrelevant features (phase component). The haze-related features are treated as negative samples, while the haze-irrelevant features serve as positive samples to provide effective contrastive feature constraints for contrastive learning. At the spatial-domain level, the reflection component is separated based on the Retinex theory and used as positive samples to guide the model in contrastive learning.

The primary contributions of this work are summarized as follows:

- (1) This study suggested a novel unsupervised image dehazing framework that integrates semantic information of both frequency and spatial domains. This method does not require paired data during the training process and has good model generalization ability.
- (2) A dual-domain convolutional mixer and a dual-domain adaptive gating fusion module are developed, which effectively enhances the feature extraction capability of the backbone network, thereby facilitating accurate depth and transmission estimation.
- (3) A multi-prior contrastive (MPC) loss is proposed, which can guide the model to preserve the critical semantic and structural features, aiming to mitigate the suboptimal model problem caused by insufficient feature constraint during unpaired training.
- (4) Numerous experiments conducted on both synthetic and real-world image datasets illustrate the proposed method performs better than current state-of-the-art models and has a good generalization capability. Also, the usefulness of our dehazing approach for infrastructure inspections is confirmed using the traffic dataset from the perspective of a UAV.

2 Related Work

2.1 Supervised Learning Dehazing Methods

Numerous supervised learning-based dehazing methods have demonstrated their efficacy in synthetic datasets, owing to the advancement of deep learning and large-scale datasets. Classic examples including MSCNN [30] and DehazeNet [31] were employed in early dehazing algorithms to learn physical characteristics such as atmospheric light and transmission maps. However, when estimating atmospheric light and transmission maps, these models are vulnerable to cumulative inaccuracies. To solve this, AODNet [32] redesigned the equation based on the atmospheric scattering model, simultaneously estimating the projection map and atmospheric light, but lacked effective prompt priors. In recent years, many supervised dehazing methods use an end-to-end methodology to directly recover haze-free images from haze photographs without taking atmospheric scattering models into account. For instance, FFANet [10] was the first to introduce the attention mechanism for image dehazing and achieved remarkable results on synthetic datasets. Babu et al. [11] proposed an effective framework validated on synthetic indoor, outdoor, and even night-time hazy images, demonstrating superior restoration quality. AECRNet [13] proposed contrastive regularization to constrain the dehazed images, effectively improving the dehazing performance. Furthermore, many supervision methods tend to design complicated modules or swap out various backbone networks such as Transformer [33–36] and Mamba [37,38] to enhance performance. The aforementioned approaches, however, rely too heavily on paired training data, which makes the model prone to overfitting during training, particularly when the generalization capacity is inadequate in practical situations [39,40].

2.2 Unsupervised Learning Dehazing Methods

In contrast to those supervised methods, unsupervised dehazing methods do not depend on paired data and can enhance image dehazing generalization in actual haze conditions. For example, Li et al. proposed an unsupervised single-image dehazing network YOLY [41], which decomposes hazy images into scene content,

transmission, and atmospheric light layers for reconstruction. However, its training process without clean images fails to capture the clean image domain's intrinsic properties, limiting dehazing performance. Zhao et al. [20] introduced a dark channel prior dehazing technique employing weakly supervised learning, albeit with substantial computational overhead. Wang et al. [18] propose an unsupervised contrastive learning paradigm based on the GAN framework, enhancing generalization in real haze scenarios, yet struggling with image detail restoration in dense haze conditions. Chen et al. [42] integrated a pre-trained dehazing model with physical priors, enabling adaptation to real-world dehazing scenarios, highlighting the efficacy of physical prior constraints in enhancing unpaired dehazing methods. Moreover, RIDCP [43] achieved notable real-world dehazing outcomes by leveraging high-quality codebook priors, albeit reliant on adjusting the haze synthesis coefficient.

To date, most unsupervised dehazing frameworks leverage generative adversarial networks (GANs) as benchmark. For instance, Engin et al. [22] proposed an end-to-end unsupervised dehazing network that enhances the CycleGAN by integrating cyclic consistency and perceptual loss. However, this method neglects the significance of depth information in the generation of haze images, which can result in the production of unrealistic haze and subsequently impair dehazing performance. Yang et al. [19] address this issue by modeling the depth map and scattering coefficient while simultaneously introducing bidirectional contrast loss to improve dehazing efficacy. Nonetheless, the absence of prior knowledge constraints can lead to inaccuracies in texture and color during the dehazing process. In our study, we incorporate frequency-domain priors and Retinex physical theories into both the dehazing and hazing processes, thereby providing effective feature constraints for the generated images.

3 Methodology

3.1 Overall Architecture

The lack of paired ground truth in unsupervised dehazing methods leads to the loss of detail and texture information. Fortunately, frequency-domain prior knowledge can assist the learning process of unsupervised dehazing frameworks [24]. To strike a balance between generalization capability and dehazing performance, this study proposes the Unpaired Dual-Domain Dehazing Network (UD³Net), the overall architecture of which is illustrated in Fig. 1.

UD³Net comprises a dehazing-hazing branch and a hazing-dehazing branch, along with multi-prior contrastive learning; both branches contain dehazing and hazing processes based on the Atmospheric Scattering Model (ASM) [27,44], as shown in Eq. (1):

$$I(x) = J(x)t(x) + A(1 - t(x)), t(x) = e^{-\beta d(x)} \quad (1)$$

where, $I(x)$ is the hazy image, $J(x)$ is the corresponding clear image, A represents the global atmospheric light, $t(x)$ is the transmission map, β is the scattering coefficient and $d(x)$ is the corresponding depth information. It is worth noting that the transmission estimator and the depth estimator share the same backbone network, integrating features from the spatial domain and the frequency domain to obtain the transmission map or depth map. Next, we will introduce these two dual-domain physical networks in detail.

In the dehazing network, as the green part in Fig. 1: given the input hazy image $I_h(x)$, the transmission estimator \mathcal{G}_t estimates the transmission map $\hat{t}(x)$ and scattering coefficient $\hat{\beta}$, thereby reconstructing clear images through atmospheric scattering models. This process can be described as follows:

$$\hat{J}_c(x) = \frac{I_h(x) - \hat{A}}{\hat{t}(x)} + \hat{A}, \hat{t}(x), \hat{\beta} = \mathcal{G}_t(I_h(x)) \quad (2)$$

where \hat{A} is the global atmospheric light estimated using the dark channel prior [4].

In the hazing network, as the blue part in Fig. 1: The depth map $\hat{d}(x)$ of a clear image $J_c(x)$ is estimated by the depth estimator \mathcal{G}_d , as shown the blue module, and then a β is provided. Combined with the depth map $\hat{d}(x)$ and the scattering coefficient β , the hazy image $\hat{I}_h(x)$ based on Eq. (1) is generated. These processes can be expressed as:

$$\hat{d}(x) = \mathcal{G}_d(J_c(x)), \hat{I}_h(x) = J_c(x)e^{-\beta\hat{d}(x)} + A(1 - e^{-\beta\hat{d}(x)}) \quad (3)$$

where, β is the scattering coefficient, in the dehazing-hazing branch, β is estimated by the dehazing network, and in the hazing-dehazing branch, β is randomly sampled within a predetermined range; in this paper, the range of β is from 0.6 to 1.8.

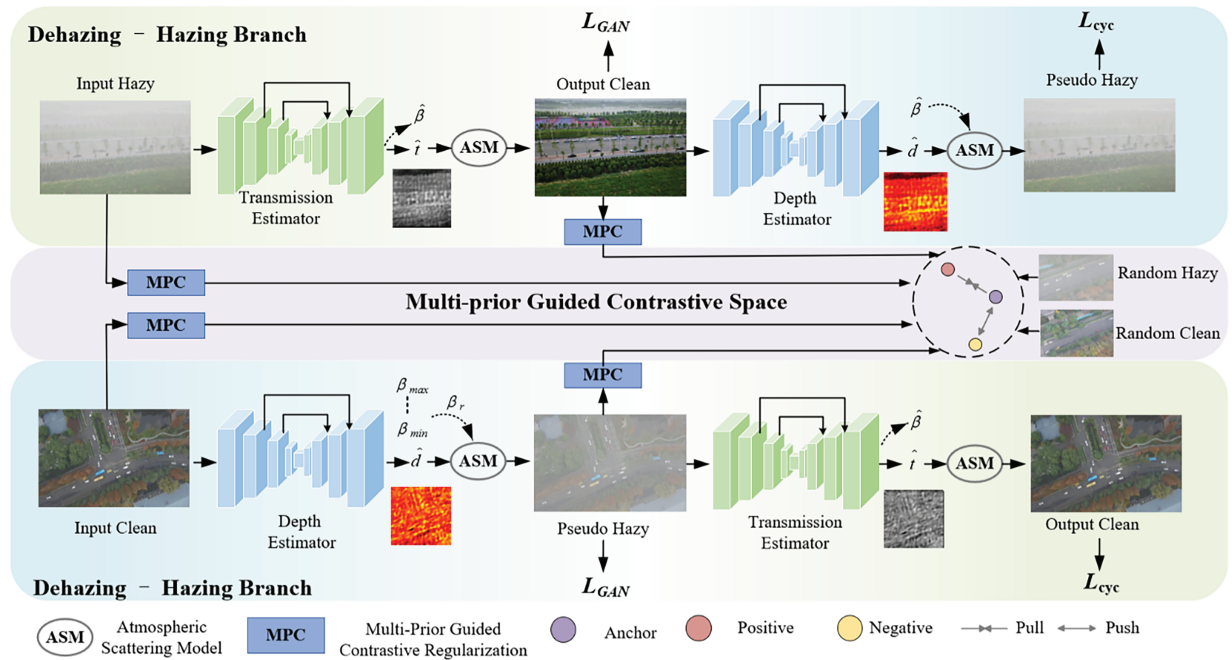


Figure 1: Overall architecture of the unpaired dual-domain dehazing network (UD³Net). During testing, only the dehazing network (green) needs to be activated.

3.2 Dual-Domain Convolution Mixer

As shown in Fig. 2, the structures of transmission estimator and depth estimator are similar to U-Net, core components which are the dual-domain convolutional mixer (DCM) and the dual-domain adaptive gated fusion (DAG). In dehazing, a hazy image is utilized as input, and the goal is to obtain the corresponding transmission map. When adding haze, the input is a clear image, and the output is a depth map of the clear image. The following sections introduce the main configurations of DCM and DAG.

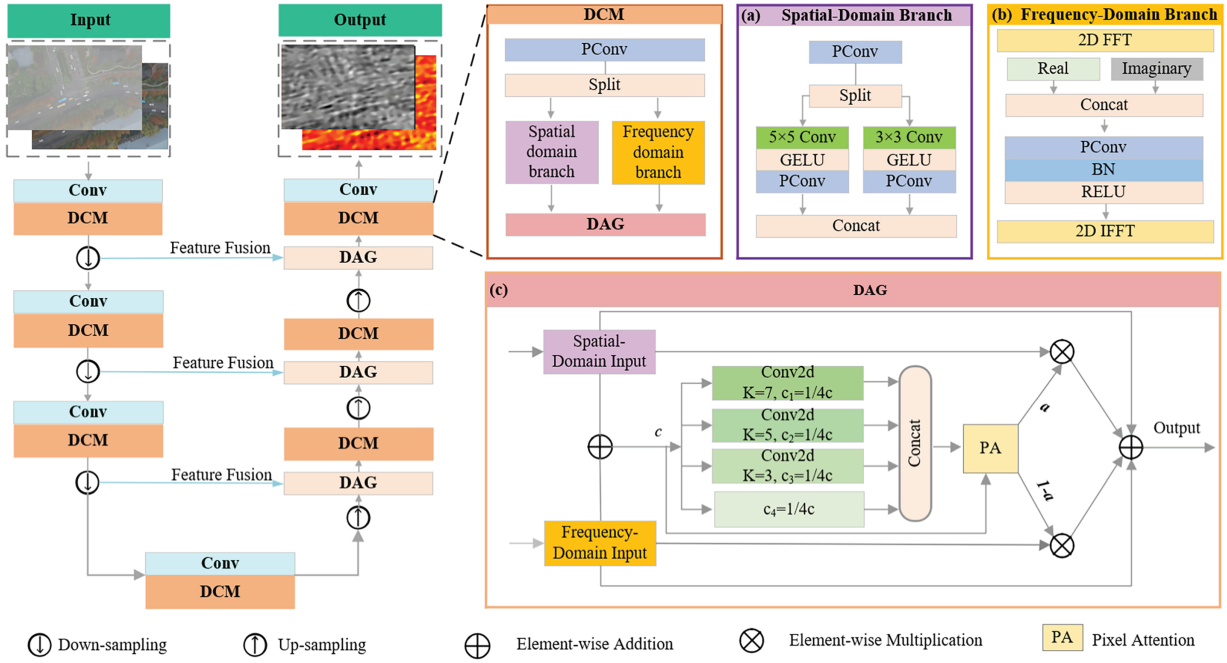


Figure 2: The overall structure of the transmission estimator and depth estimator mainly consists of three parts: encoder, decoder and feature fusion.

Traditional Convolutional Neural Networks (CNNs) process images in the spatial domain, which is excellent for capturing local features like edges and textures. However, they often struggle with global properties, such as overall illumination and color shifts caused by haze. The DCM module addresses this by operating simultaneously in both the spatial and frequency domains.

While the spatial branch extracts local high-frequency details, the frequency branch (via FFT) efficiently captures global contextual information and low-frequency components. By mixing these two representations, DCM ensures that the network recovers fine details without losing sight of the global image structure and color consistency.

As shown in Fig. 2a, the spatial domain branch splits the input features F_s into F_s^1 and F_s^2 , and employs 3×3 convolution and 5×5 convolution, respectively, to extract multi-scale features. Lastly, to get the output spatial features, concatenate the features from the two data streams. In real-world applications, depthwise separable convolution $DWConv$ is adopted to increase efficiency. The following is a description of this process:

$$F_s^{out} = Concatenate [DWConv (F_s^1), DWConv (F_s^2)] \quad (4)$$

where, F_s^{out} represents the output spatial feature, F_s^1 is the separated feature one, F_s^2 is the separated feature two, and $DWConv$ indicates the depthwise separable convolution.

The frequency domain branch is shown in Fig. 2b. Specifically, the fast Fourier Transform (FFT) converts an image $X \in R^{H \times W \times C}$ into frequency space to produce the complex component $\mathcal{F}(u, v)$, which has the following expression:

$$\mathcal{F}(u, v) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} X(h, w) e^{-j2\pi(\frac{hu}{H} + \frac{wv}{W})} \quad (5)$$

where, u and v are the coordinates in Fourier space. In the frequency-domain branch, Ff undergoes the FFT to convert it into real and imaginary components. After aggregating the real and imaginary components in the channel dimension, 1×1 convolution is used to extract frequency-domain features. After modulation, the real and imaginary components are separated, and the Inverse Fast Fourier Transform (IFFT) converts the frequency-domain features back to the spatial domain:

$$X_{\mathcal{R}}, X_{\mathcal{I}} = \mathcal{FFT}(Ff) \quad (6)$$

$$\widehat{X}_{\mathcal{R}}, \widehat{X}_{\mathcal{I}} = \sigma \cdot BN(PCConv(Concat(X_{\mathcal{R}}, X_{\mathcal{I}}))) \quad (7)$$

$$F_f^{out} = IFFT(\widehat{X}_{\mathcal{R}}, \widehat{X}_{\mathcal{I}}) \quad (8)$$

where, $X_{\mathcal{R}}$ represents the real part, $X_{\mathcal{I}}$ represents the imaginary part, F_f denotes the input to the frequency domain branch, and F_f^{out} represents the output of the frequency domain branch.

Simply concatenating spatial and frequency features is suboptimal because their contribution varies across different image regions. Additionally, we have found that combining the features of the encoder and decoder parts is a useful method for dehazing and other underlying visual tasks [45,46] in an architecture akin to U-Net. In order to restore photos free of haze, low-level features are essential. However, when combining low-level and high-level characteristics using the element-wise addition method [9,45,47], there is an issue with mismatch in receptive fields [48]. To address the aforementioned challenges, this study proposes a novel feature fusion method (denoted as DAG) that enables adaptive feature fusion. As shown in Fig. 2c, it dynamically learns a weight map to determine the proportion of information to be retained from each domain at a pixel level. This allows the network to selectively fuse the most distinct features from both domains while suppressing noise or redundant information. This method can be described as follows:

$$F_{fuse} = F_s^{out} \oplus F_f^{out} \quad (9)$$

$$F_m = Concat(Conv_{3 \times 3}(F_{fuse}), Conv_{5 \times 5}(F_{fuse}), Conv_{7 \times 7}(F_{fuse}), F_{fuse}) \quad (10)$$

where, F_{fuse} represents the addition of two-domain features, and F_m represents the output of multi-scale convolution. Inspired by the success of multi-scale receptive fields in recent dehazing advancements [49], we implement parallel convolutions with kernel sizes of 3×3 , 5×5 , and 7×7 . The 3×3 kernels focus on extracting local structural details, whereas the larger 5×5 and 7×7 kernels are responsible for capturing broader context to handle uneven haze distributions. This cooperation completes the initial dual-domain feature aggregation. Subsequently, the pixel attention is used to calculate the gating weights for the fused features F_m and the attention map:

$$a = Sigmoid(PA(F_m, F_{fuse})) \quad (11)$$

where, a represents the gating factor, which controls the contribution of spatial domain features and frequency domain features through the attention mechanism, the final output feature F_{DAG} is calculated in the following way:

$$F_{DAG} = Conv_{1 \times 1}(F_s^{out} + F_f^{out} + a \cdot F_s^{out} + (1 - a) \cdot F_f^{out}) \quad (12)$$

3.3 Multi-Prior Contrastive Learning

Motivated by frequency-domain prior and the Retinex theory, we have created a multi-prior contrastive (MPC) loss to address the drawback of feature constraints in traditional contrastive learning.

(1) Retinex prior

The Retinex theory [50], a classical computational model of color constancy in the human visual system, posits that the image perceived by the human eye, I , can be decomposed into two multiplicative components: a slowly varying illumination component L and a rapidly changing reflection component R , expressed as: $I = L \cdot R$, where the image I can be decomposed into the illumination component L and the reflection component R , and \cdot represents the product of the elements.

The illumination component L describes the lighting conditions incident on the scene, while the reflection component R represents the intrinsic properties of objects and is the fundamental carrier of image details and color [51]. Viewing the image dehazing problem through this theoretical framework, it becomes evident that haze, acting as a medium of suspended atmospheric particles, primarily interferes with the path and intensity of light reaching the imaging sensor through absorption and scattering. This interference does not alter the essential reflective properties of the scene objects; instead, it is integrated into the illumination component L as a form of degradative environmental illumination L . Consequently, the formation of a hazy image can be regarded as the result of the clear scene's reflection component being imaged under an illumination field corrupted by haze. This naturally leads to a key prior constraint: in an ideal dehazing process, the dehazed clear image and the original hazy image should share the same reflection component R . The formula is expressed as:

$$I_h = L_h \cdot R, I_c = L_c \cdot R \quad (13)$$

where, I_h represents the haze image, I_c represents the clear image corresponding to I_h , L_h and L_c respectively represent the illumination components of the haze image and the clear image, and R represents their common reflection component.

(2) Frequency-domain prior

Numerous earlier investigations [28,52] have demonstrated that the degradation of haze is mainly reflected in the amplitude spectrum of the image spectrum. As shown in Fig. 3, when the amplitude spectra of the hazy image and the clean image remain unchanged and the phase spectra are exchanged (as shown in Fig. 3a,c), the resulting image is very similar in clarity to the source image. When the amplitude spectrum is exchanged while keeping the phase spectrum unchanged (as shown in Fig. 3b,d), the clarity of the obtained image differs significantly from that of the source image. Furthermore, swapping the phase spectra will alter the background information of the resulting images when the two photos have inconsistent backgrounds (as shown in Fig. 3c,d). As can be seen above, the phase spectrum mostly contains the image's background information, but the amplitude spectrum primarily reflects the haze's deterioration characteristics.

In unsupervised dehazing networks, ground truth is unavailable. If only randomly selected images are used to construct contrastive learning samples in the spatial domain, it is prone to result in the problem of weak feature constraints. To address this challenge, this paper proposes a multi-prior contrastive loss (MPC), whose structure is illustrated in Fig. 4. By fully leveraging abundant image prior knowledge to guide the contrastive learning process, this module effectively mitigates the aforementioned issue of insufficient constraints. Its core mechanism is as follows: in the feature space, the MPC encourages anchors to approach positive samples while moving away from negative samples. As such, both the dehazing network and the hazing network can benefit from the constraints imposed by the MPC. For example, during the dehazing process, the MPC constrains the UD³Net to align the haze-irrelevant components (i.e., phase spectrum and

reflectance) of the generated dehazed image with those of the corresponding hazy image, while differentiating their haze-relevant components (i.e., amplitude spectrum). Given the hazy image input I_h and the dehazed image J_c , the output amplitude spectrum, phase spectrum and reflection component can be described as follows:

$$A_{u,v}(I_h), P_{u,v}(I_h) = DFT(I_h) \quad (14)$$

where, $A_{u,v}$ and $P_{u,v}$ represent the phase spectrum and amplitude spectrum respectively, (u,v) represents the coordinate in the frequency domain, and DFT stands for Discrete Fourier Transform.

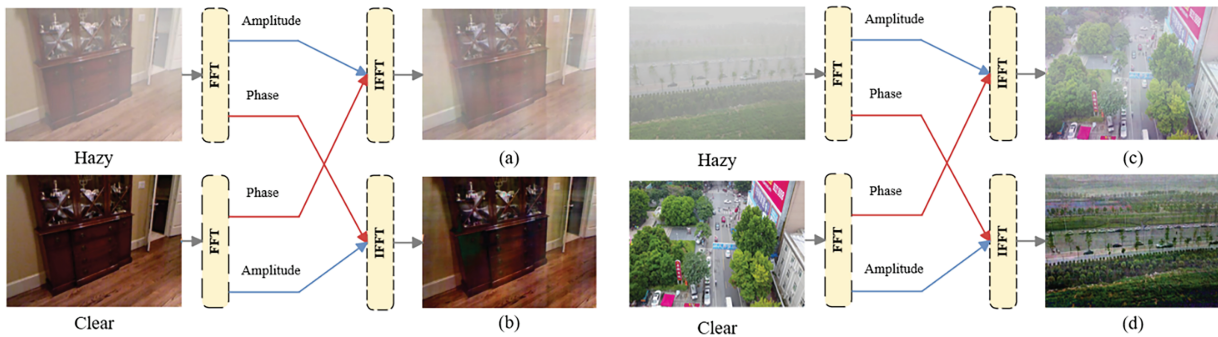


Figure 3: Visual analysis of the relationship between haze degradation and amplitude spectrum and phase spectrum characteristics in the frequency domain. The FFT represents the fast Fourier transform, and the IFFT represents the inverse Fast Fourier transform. We denote the image with hazy image amplitude and clear image phase as (a), and the image with clear image amplitude and hazy image phase as (b). (c,d) are similar.

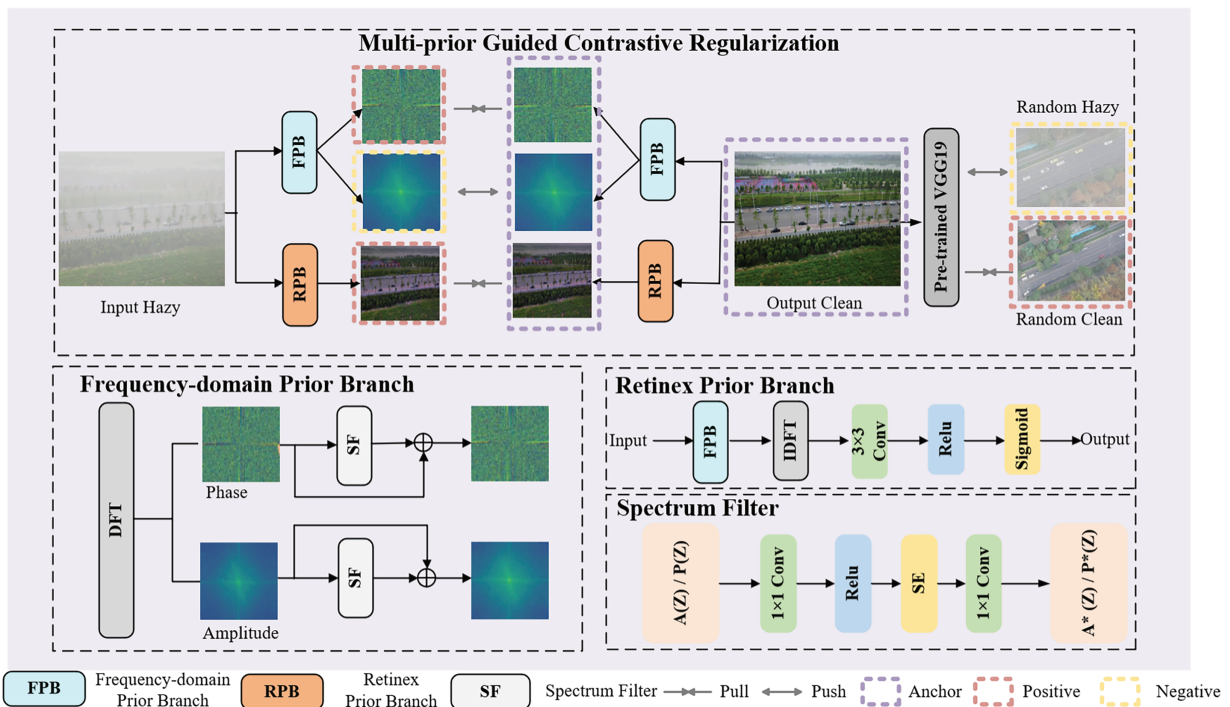


Figure 4: The multi-prior contrastive regularization (MPC) processing flow.

For the glare and noise in the phase spectrum and amplitude spectrum, they can be processed in the frequency domain through dynamic spectral filtering (SF) to avoid interfering with the subsequent dehazing learning. This process can be described as follows:

$$\begin{aligned} A^* u, v (I_h) &= Conv_{1 \times 1} SE (Conv_{1 \times 1} (A u, v (I_h))) \\ P^* u, v (I_h) &= Conv_{1 \times 1} SE (Conv_{1 \times 1} (P u, v (I_h))) \end{aligned} \quad (15)$$

where, SE stands for Squeeze-and-Excitation Attention, which dynamically suppresses noise with inconsistent frequency characteristics by computing channel weight maps. $A^* u, v (I_h)$ and $P^* u, v (I_h)$ represents the component after SF. Finally, the filtered components are obtained through residual connection, and the image is decoupled into degradation information (amplitude spectrum) and background information (phase spectrum).

$$\begin{aligned} \tilde{A}_{u,v} (I_h) &= A^*_{u,v} (I_h) + A_{u,v} (I_h) \\ \tilde{P}_{u,v} (I_h) &= P^*_{u,v} (I_h) + P_{u,v} (I_h) \end{aligned} \quad (16)$$

where, $\tilde{A}_{u,v} (I_h)$ and $\tilde{P}_{u,v} (I_h)$ respectively represent the outputs of the frequency-domain prior branch (FPB). Similarly, the amplitude spectrum $\tilde{A}_{u,v} (J_c)$ and phase spectrum $\tilde{P}_{u,v} (J_c)$ of the dehazed image can be obtained. In order to estimate the reflection component, the adaptive prior is automatically and iteratively learned from the hazy image and the restored image pair through a convolution operation after the output frequency-domain component has been converted back to the spatial domain using the inverse discrete Fourier transform. The following is a description of this process:

$$I'_h = IDFT (\tilde{A}_{u,v} (I_h), \tilde{P}_{u,v} (I_h)) \quad (17)$$

where, I'_h represents the feature after converting $\tilde{A}_{u,v} (I_h)$ and $\tilde{P}_{u,v} (I_h)$ back to the spatial domain, and IDFT represents the inverse discrete Fourier transform.

$$RI_h = Sigmoid (Relu (Conv_{3 \times 3} (I'_h))) \quad (18)$$

where, the reflection component of the RI_h generation output through the retinex prior branch (RPB). Similarly, the component RJ_c of the J_c after RPB processing can be obtained.

3.4 Loss Function

Using an efficient loss function to oversee model training is essential, particularly for unpaired dehazing frameworks. We constrain the learning of density and depth using pseudo-supervised parameter loss in addition to the adversarial loss and cycle consistency loss. Furthermore, a multi-prior contrastive loss is proposed to tackle the problem of inadequate feature restrictions.

- (1) **Cycle Consistency Loss:** Cycle consistency loss is a commonly used loss function in CycleGAN. In the UD³Net framework, the generated hazy images and clear images should maintain consistency with their corresponding input hazy images and input clear images. The cycle consistency loss is defined as follows:

$$\mathcal{L}_{cyc} = \mathbb{E}_{I \sim \mathcal{X}_h} [\|\mathcal{G}_h (\mathcal{G}_d (I)) - I\|_1] + \mathbb{E}_{J \sim \mathcal{X}_c} [\|\mathcal{G}_d (\mathcal{G}_h (J)) - J\|_1] \quad (19)$$

where, I stands for the original hazy image and J for the original clean image. The dehazing and hazing networks in the two branches are denoted by $\mathcal{G}_d (I)$ and $\mathcal{G}_h (J)$, respectively. $\|\cdot\|_1$ denotes the $\mathcal{L}1$ norm, \mathbb{E} represents the mathematical expectation.

- (2) Adversarial Loss: The images generated by adversarial loss constrained dehazing and hazing networks are visually more realistic. The LSGAN [53] we employ can be expressed as:

$$\begin{aligned}\mathcal{L}_{adv}(\mathcal{D}_c) &= \mathbb{E} \left[(\mathcal{D}_c(\hat{J}))^2 \right] + \mathbb{E} \left[(\mathcal{D}_c(J) - 1)^2 \right] \\ \mathcal{L}_{adv}(\mathcal{G}_d) &= \mathbb{E} \left[(\mathcal{D}_c(\hat{J}) - 1)^2 \right]\end{aligned}\quad (20)$$

where, \mathcal{D}_c denotes the discriminator used to determine whether the input image belongs to the clean domain, \hat{J} represents the dehazing result from \mathcal{G}_d , and J denotes the clean image. For the hazing network \mathcal{G}_h and its corresponding discriminator \mathcal{D}_h , the adversarial loss adopts the same form.

- (3) Pseudo-Supervised Parameter Loss: To address the challenge of lacking ground truth for depth and scattering coefficients, we introduce the pseudo-supervised parameter loss \mathcal{L}_{psp} . According to Eq. (1), we compute a pseudo-ground truth depth d from the hazy image to constrain the estimated depth \hat{d} in Eq. (3). The randomly sampled scattering coefficient β constrains the estimated scattering coefficient $\hat{\beta}$ in Eq. (2). The final pseudo-supervised parameter loss is expressed as follows:

$$\begin{aligned}\mathcal{L}_d &= \|\hat{d} - d\|_1 \\ \mathcal{L}_\beta &= \|\hat{\beta} - \beta\|_2 \\ \mathcal{L}_{psp} &= \mathcal{L}_d + \mathcal{L}_\beta\end{aligned}\quad (21)$$

where, $\|\cdot\|_2$ denotes the $\mathcal{L}2$ norm.

- (4) MPC Loss: Prior unsupervised dehazing techniques [13,19] ignore exploitable information in the original image. The proposed multi-prior contrastive loss (MPC) leverages prior knowledge to provide supervisory information from the original image, pulling the restored image closer to positive samples while pushing it farther from negative samples. The selection of positive and negative samples is detailed in Fig. 4. MPC can be formulated as:

$$\mathcal{L}_{mpc} = \sum_{i=1}^n \omega_i \frac{\|V_i(\hat{c}) - V_i(J)\|_1}{\|V_i(\hat{c}) - V_i(I)\|_1} + \frac{\|P - \tilde{P}\|_2}{\|A - \tilde{A}\|_2} + \|R - \tilde{R}\|_2 \quad (22)$$

where, V_i represents the features extracted from the i^{th} layer of VGG19 [54], and ω_i denotes the weights of the i th layer. Following the established settings in previous dehazing literature [18,19], we utilized features extracted from layers 3, 5, and 13 of the pre-trained VGG19 model, with corresponding weights assigned as 0.4, 0.6, and 1. \hat{c} represents the output of the dehazing network. A and P denote the amplitude spectrum and phase spectrum of \hat{c} , respectively, while \tilde{A} and \tilde{P} denote the amplitude spectrum and phase spectrum of the input hazy image. R and \tilde{R} represent the reflected component of \hat{c} and the reflected component of the input hazy image, respectively. Similarly, the hazing network also benefits from MPC.

- (5) Total loss: Ultimately, the total loss function \mathcal{L}_{total} of the UD³Net framework can be expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{cyc} + \lambda_{adv} \mathcal{L}_{adv} + \mathcal{L}_{psp} + \lambda_{mpc} \mathcal{L}_{mpc} \quad (23)$$

where, λ_{adv} and λ_{mpc} are the weights for balancing different loss terms. We experimentally set λ_{adv} and λ_{mpc} to 0.2 and 0.0001, respectively.

4 Experiments

4.1 Implementation Details

- (1) **Dataset:** The RESIDE dataset [55] serves as a prominent benchmark for dehazing research. Initially, we conducted training using the ITS training set from the RESIDE dataset, which comprises 13,990 synthetic haze images captured indoors, along with their corresponding reference images. Subsequently, we utilized four haze datasets as test sets to assess the model's performance. The SOTS-indoor [55] and SOTS-outdoor [55] test sets each contain 500 haze images, representing indoor and outdoor scene synthesis datasets, respectively. The HSTS [55] dataset includes 10 synthetic outdoor haze images and 10 real-world haze images. HazyDet-test [56] features aerial photography images captured from a drone's perspective, consisting of 2000 pairs of haze images synthesized based on depth information. Additionally, the model was trained using unpaired outdoor real haze images, which include 3577 clean outdoor images and 2903 high-quality haze images sourced from RESIDE, and was evaluated on the RTTS dataset [55]. The RTTS dataset contains 4322 real outdoor haze images and lacks corresponding reference images.
- (2) **Evaluation metrics:** For the paired test set, we employ two widely recognized full-reference metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), to evaluate the performance of the dehazing method. In the case of the RTTS dataset, where ground truth is unavailable, we utilize three no-reference evaluation metrics, FADE [57], BRISQUE [58], and NIQE [59], for comparative analysis. FADE is utilized to assess the haze density within a given image, whereas BRISQUE and NIQE evaluate the overall quality of the image.
- (3) **Compared methods:** To assess the effectiveness of UD³Net, we selected several representative dehazing algorithms for comparison. This selection includes supervised dehazing methods such as AODNet [32], FFANet [10], PSD [42], AEARNet [13], and C2PNet [14], as well as unsupervised methods including DCP [4], CycleGAN [16], RefineDNet [20], YOLY [41], D4 [17], POGAN [24], UCL [18], and D4+ [19]. To ensure fairness, the dehazing methods made public in the code use the same training data. POGAN provided the results cited in the paper. We also contrasted it with some representative techniques in actual dehazing circumstances, such as PSD [42] and RIDCP [43], while testing on the RTTS dataset.
- (4) **Implementation Details:** All experiments were conducted using NVIDIA RTX 3080 Ti GPUs. During the training phase, UD³Net was trained for 1.5×10^5 iterations with a batch size of 2. The learning rate was established at 0.0001 based on prior experience. Training images were randomly cropped to 256×256 pixels, and data augmentation was implemented through horizontal flipping.

4.2 Ablation Experiment

Ablation studies on UD³Net. To validate the effectiveness of the DCM module and the MPC within UD³Net, we conducted ablation experiments on the Outdoor dataset. Four experimental groups were designed: (a) UD³Net: basic unsupervised dehazing framework; (b) UD³Net +DCM; (c) UD³Net +MPC; (d) UD³Net +DCM +MPC. Performance summaries for each model are presented in Table 1. The results show that integrating DCM can increase PSNR by 0.48 dB and SSIM by 0.2% at the same time. The integration of MPC increased PSNR by 0.69 dB and SSIM by 0.6%. When both DCM and MPC are applied concurrently, PSNR reaches 26.45 dB and SSIM increases to 96.5%, surpassing the UD³Net by 1.61 dB and 1.5%, respectively. The ablation curves are shown in Fig. 5a,d.

Table 1: Ablation study of our method on the SOTS-outdoor dataset.

Method	DCM	MPC	PSNR \uparrow	SSIM \uparrow
a			24.84	0.950
b		✓	25.53	0.956
c	✓		25.32	0.952
d	✓	✓	26.45	0.965

Note: (a) refers to the original UD³Net, (b) UD³Net + DCM, (c) UD³Net + MPC, and (d) UD³Net + DCM + MPC. The symbol \uparrow indicates that higher values correspond to better quality. The best performance will be highlighted in bold.

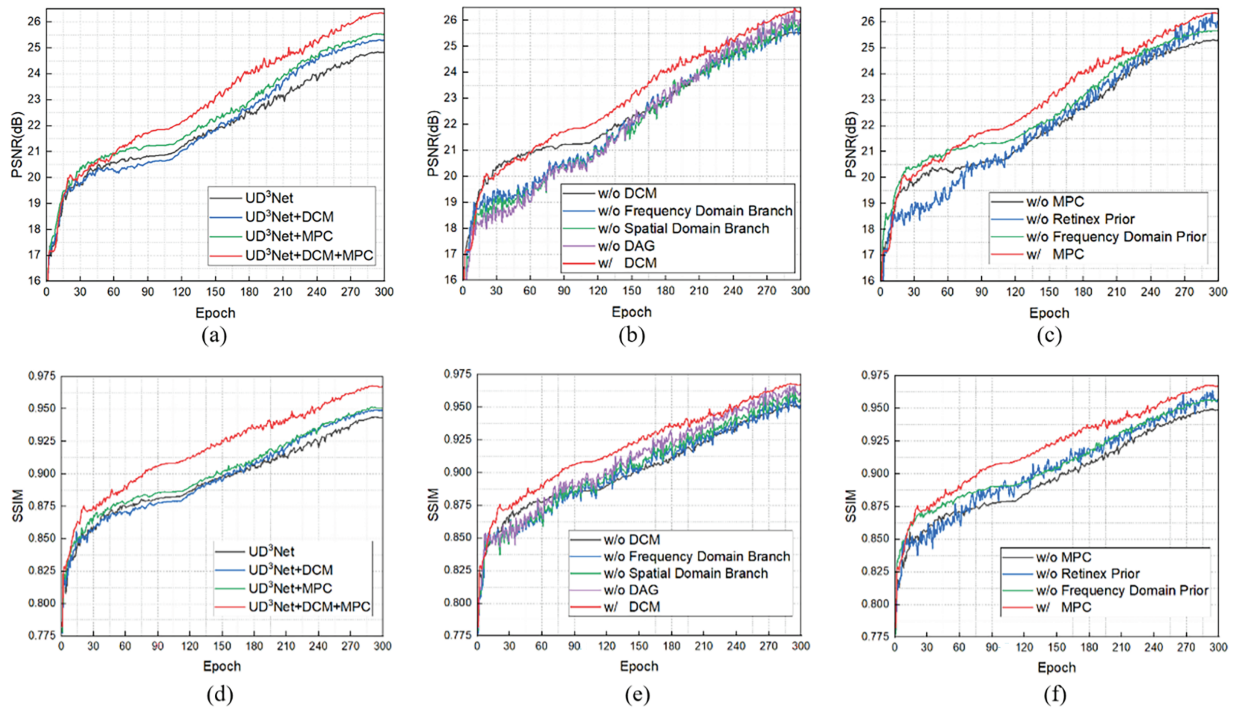


Figure 5: PSNR and SSIM curves of the proposed method on the SOTS-outdoor dataset. (a) It is a PSNR curve based on UD³Net, (b) is a PSNR curve based on DCM, (c) is a PSNR curve based on MPC, likewise, (d–f) are SSIM curves.

Ablation studies on DCM. To further investigate the effectiveness of DCM, we progressively removed the spatial domain branch, frequency domain branch, and DAG from DCM. As shown in Table 2, the optimal results were achieved when using the complete DCM. The results indicate that replacing the proposed DAG with a simple Concatenation operation leads to a performance drop of 0.36 dB in PSNR. The theoretical advantage lies in the dynamic feature selection capability of DAG. In contrast, DAG employs an attention-based gating mechanism. By computing spatial and channel-wise attention maps, DAG explicitly re-weights the incoming features. This allows the network to adaptively emphasize useful information. The ablation curves are depicted in Fig. 5b,e.

Ablation studies on MPC. To further investigate the effectiveness of MPC, we progressively removed the frequency-domain prior and retinex prior from MPC, as shown in Table 3. Both the absence of prior guidance and the use of a single prior resulted in significant performance degradation. Fig. 5c,f illustrates the ablation curves corresponding to our experiments.

Table 2: Ablation study of DCM on the SOTS-outdoor dataset.

DAG	FDB	SDB	PSNR↑	SSIM↑
			25.53	0.956
		✓	25.80	0.957
	✓		25.96	0.959
	✓	✓	26.09	0.962
✓	✓	✓	26.45	0.965

Note: FDB and SDB correspond to the frequency domain branch and spatial domain branch in DCM, respectively. The symbol ↑ indicates that higher values correspond to better quality. The best performance will be highlighted in bold.

Table 3: Ablation study of MPC on the SOTS-outdoor dataset.

Frequency Domain Prior	Retinex Prior	PSNR↑	SSIM↑
		25.32	0.952
✓		26.09	0.962
	✓	25.64	0.956
✓	✓	26.45	0.965

Note: The symbol ↑ indicates that higher values correspond to better quality. The best performance will be highlighted in bold.

Fig. 5 presents the PSNR and SSIM curves corresponding to the ablation experiment. Specifically, the PSNR and SSIM curves of the proposed UD³Net are illustrated in Fig. 5a,d. As the number of epochs increases, the proposed method (red curve) achieves the highest PSNR and SSIM values. This superiority can be attributed to the multi-scale feature extraction capability of the DCM module and the effective feature constraint role of the MPC. Furthermore, Fig. 5b,e substantiate the contributions of the constituent components within the DCM module, while Fig. 5c,f validate those of the individual prior constraints in the MPC.

Results derived from the sensitivity analysis of loss weights are summarized in Table 4. We varied the value of λ_{mpc} (e.g., 0.00001, 0.0001, 0.001, 0.01) while keeping other parameters fixed. The results show that setting λ_{mpc} to 0.0001 achieves the best balance. Lower values weaken the guidance of priors, while excessively high values lead to unstable training dynamics. Similarly, for the adversarial weight λ_{adv} , the experimental test values range from 0.1 to 0.4, increasing by 0.1 each time. The results indicate that increasing or decreasing the value of λ_{adv} from 0.2 leads to a slight decrease in dehazing performance.

Table 4: Ablation study of loss function weights on the SOTS-outdoor dataset.

Weights	PSNR	SSIM
$\lambda_{adv} = 0.1$	24.91	0.952
$\lambda_{adv} = 0.2$	26.45	0.965
$\lambda_{adv} = 0.3$	25.32	0.954
$\lambda_{adv} = 0.4$	25.18	0.950
$\lambda_{mpc} = 0.00001$	25.42	0.954
$\lambda_{mpc} = 0.0001$	26.45	0.965

(Continued)

Table 4 (continued)

Weights	PSNR	SSIM
$\lambda_{\text{mpc}} = 0.001$	25.90	0.957
$\lambda_{\text{mpc}} = 0.01$	25.73	0.957

4.3 Comparison Experimental

To assess the generalization capability and performance of UD³Net across various haze scenarios, it was initially compared with state-of-the-art dehaze techniques using three benchmark datasets. Notably, all supervised methods utilized pre-trained weights from the ITS dataset. In contrast, unsupervised methods, including UD³Net, omitted all paired information during training.

The quantitative results are presented in Table 5. Analysis of the SOTS-indoor dataset reveals that C2PNet and AECRNet demonstrate exceptional performance, indicating their robust fitting capabilities. However, their performance significantly declines on the SOTS-outdoor and HSTS datasets, highlighting the overfitting issue associated with supervised methods. In contrast, UD³Net not only demonstrates exceptional performance on the SOTS-indoor dataset but also surpasses all paired and unpaired methods in terms of PSNR and SSIM on the SOTS-outdoor dataset. Additionally, it attains suboptimal PSNR and optimal SSIM on the HSTS dataset. These findings substantiate the generalization ability of UD³Net in addressing diverse haze distributions. The qualitative analysis results are shown in Figs. 6 and 7, which facilitate visual analysis.

Table 5: Quantitative comparisons with paired and unpaired dehazing methods were conducted on the SOTS-indoor, SOTS-outdoor, and HSTS datasets.

Method	SOTS-Indoor		SOTS-Outdoor		HSTS		
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
Paired	AODNet	19.06	0.850	19.63	0.892	20.55	0.897
	FFANet	36.36	0.993	20.23	0.905	20.88	0.904
	PSD	15.02	0.764	15.63	0.834	15.55	0.822
	AECRNet	37.17	0.990	19.53	0.875	15.32	0.818
	C2PNet	42.56	0.995	16.84	0.810	14.70	0.755
Unpaired	DCP	13.10	0.699	20.15	0.919	14.84	0.761
	CycleGAN	21.34	0.899	20.55	0.856	18.52	0.831
	RefineDNet	24.36	0.939	19.84	0.853	21.69	0.904
	YOLY	15.84	0.819	14.75	0.857	23.82	0.913
	D4	25.42	0.932	25.83	0.956	25.03	0.957
	POGAN	25.51	0.934	25.90	0.956	24.11	0.917
	UCL	20.05	0.836-	25.21	0.927	26.87	0.933
	D4+	25.79	0.937	26.30	0.960	24.43	0.903
	Ours	26.03	0.942	26.45	0.965	25.23	0.958

Note: The symbol ↑ indicates that higher values correspond to better quality. ‘-’ denotes unavailable values. Bold text denotes the best results.

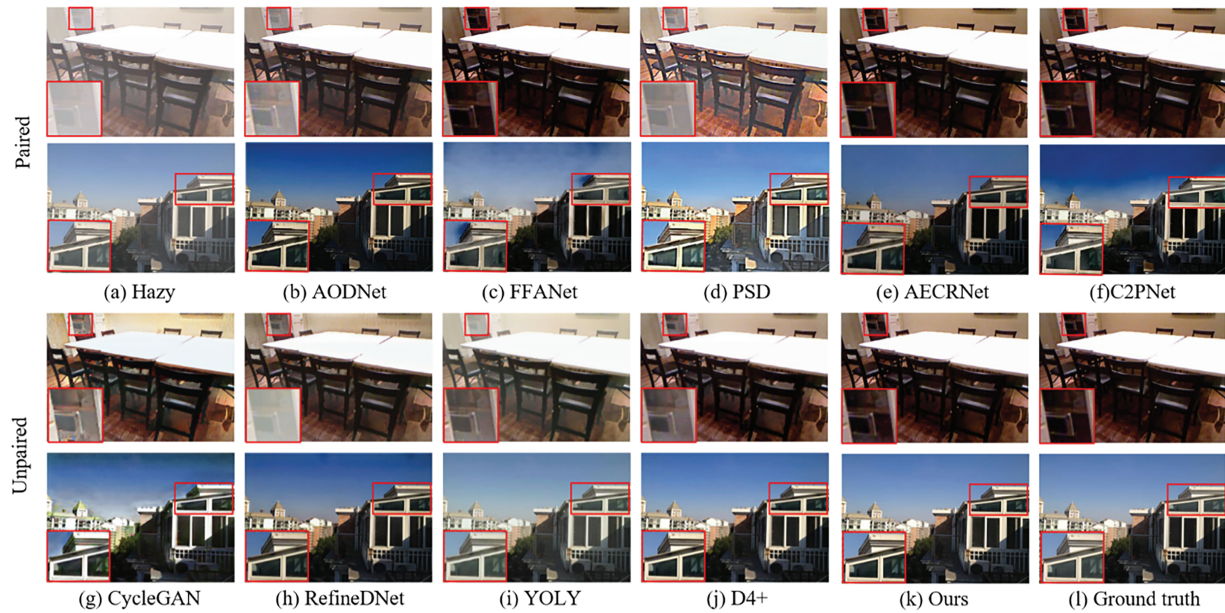


Figure 6: Qualitative comparison of different image dehazing methods on the SOTS-indoor and SOTS-outdoor datasets.

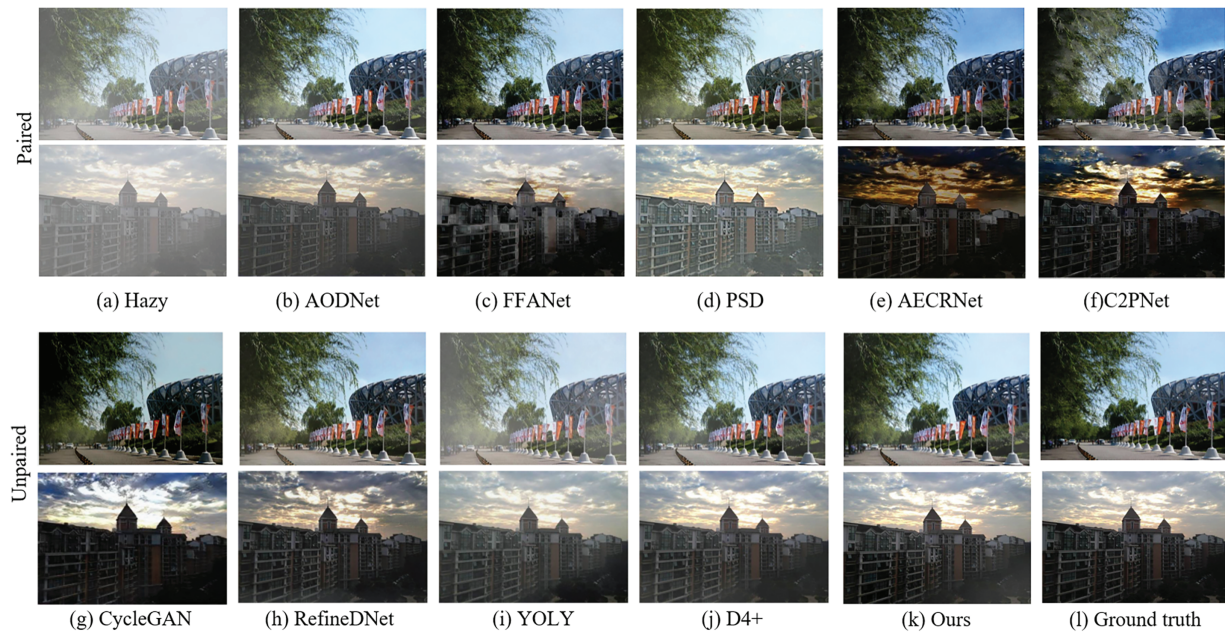


Figure 7: Qualitative comparison of different image dehazing methods on the HSTS datasets.

Fig. 6 is a visual comparison on the SOTS-indoor and SOTS-outdoor datasets. The results illustrate that most supervised methods trained on the ITS dataset exhibit challenges in generalizing to outdoor scenes. The results from FFANet and C2PNet demonstrate significant color distortion in the sky region, as shown in Fig. 6c,f, while the dehazing outcomes of AODNet, PSD, and AECRNet suffer from a loss of texture details, as illustrated in Fig. 6b,d,e. On the other hand, among the unsupervised methods, CycleGAN and

RefineDNet are susceptible to artifacts, as presented in Fig. 6g,h, and both YOLY and D4+ display minor haze residue, as displayed in Fig. 6i,j. In comparison to other methods, the dehazing result of UD³Net is more akin to a clean image, effectively mitigating color deviation and detail loss, as shown in Fig. 6k. In Fig. 7, FFANet exhibits significant artifacts, as presented in Fig. 7c, while the results produced by AECRNet and C2PNet generally demonstrate higher contrast, as illustrated in Fig. 7e,f. Additionally, the residual haze present in the outputs of other methods exceeds that observed in UD³Net, demonstrating the effectiveness of prior constraints in unpaired unsupervised learning.

To verify the performance of UD³Net on haze images from the perspective of unmanned aerial vehicles, comparative experiments were conducted on the HazyDet-test dataset. Our approach employed the ITS dataset as the training set, whereas other dehazing methods utilized pre-trained weights provided by the author. Additionally, to further evaluate the efficacy of UD³Net in authentic haze conditions, we performed tests on the RTTS dataset. In real outdoor scenarios, our method was trained on unpaired outdoor haze images.

Benefiting from the proposed framework, which allows for training on unpaired real-world datasets, UD³Net effectively bridges the domain gap between synthetic and real-world scenarios, demonstrating remarkable generalization capabilities. Furthermore, the MPC explicitly constrains the haze-irrelevant information of the output to remain consistent with the input. This mechanism ensures that the network preserves semantic content while enhancing visibility, regardless of variations in haze density. Consequently, this leads to improved robustness across diverse haze conditions. Table 6 shows that the proposed UD³Net outperforms all paired and unpaired dehazing methods on the HazyDet test dataset. On the RTTS dataset, it attains the best FADE and BRISQUE scores, and despite a slightly lower NIQE score than PSD, it delivers higher fidelity of the restored images. Visual results are illustrated in Figs. 8 and 9.

Table 6: Quantitative comparisons with paired and unpaired dehazing methods conducted on the UAV-based dataset HazyDet-test and the real hazy image dataset RTTS.

Method	Hazydet-Test		RTTS			
	PSNR↑	SSIM↑	FADE↓	BRISQUE↓	NIQE↓	
Paired	AODNet	16.69	0.799	1.023	32.064	4.764
	FFANet	19.25	0.789	2.086	33.188	4.930
	PSD	14.21	0.675	0.920	25.239	3.874
	AECRNet	–	–	1.715	29.011	4.835
	C2PNet	19.31	0.832	2.073	34.277	5.035
Unpaired	DCP	16.98	0.824	0.932	32.448	4.534
	CycleGAN	–	–	1.133	26.608	4.349
	RefineDNet	15.30	0.776	0.970	18.912	4.215
	D4	18.67	0.827	1.358	33.206	4.521
	RIDCP	16.15	0.718	0.944	18.782	4.156
	D4+	20.82	0.858	0.955	32.640	4.339
	Ours	21.17	0.866	0.856	18.152	4.206

Note: The symbol ↑ (↓) indicates that higher (lower) values correspond to better quality. ‘–’ denotes unavailable values. Bold text represents the best results.



Figure 8: Qualitative comparison of different image dehazing methods on the HazyDet-test datasets.



Figure 9: Qualitative comparison of different image dehazing methods on the real-world haze RTTS datasets.

As illustrated in Fig. 8, the aerial images obtained from the unmanned aerial vehicle exhibit a non-uniform distribution. Most supervised dehazing methods are ineffective in removing the haze present in these images. While PSD improves clarity, the resulting dehazed images exhibit excessively high saturation,

as presented in Fig. 8d. Among the unsupervised dehazing techniques, RefineDNet produced significant distortion in distant hazy regions, as shown in Fig. 8f, whereas the other methods left substantial areas of haze. In contrast, UD³Net demonstrates superior performance in haze removal from the drone's perspective, retaining only a minimal amount of residual haze, as displayed in Fig. 8i.

Fig. 9 shows visual comparison of different dehazing models on real-world datasets. The first line illustrates the limitations of the advanced supervised learning methods in dehazing. It is worth noting that while PSD achieves lower NIQE scores in Table 6, a closer inspection of Fig. 9d reveals that these scores are largely driven by excessive contrast enhancement. Due to its sensitivity to local contrast statistics, NIQE may favor over-sharpened features and does not reliably reflect the presence of unnatural artifacts caused by domain shift. Additionally, in unsupervised learning methods, both CycleGAN and RefineDNet, exhibit significant distortion in their dehazing outcomes, as shown in Fig. 9f,g. Note although the results from D4 and D4+ appear visually natural, they suffer from color shifts and retain a considerable amount of haze, as depicted in Fig. 9h,i. In comparison, the proposed UD³Net, with its robust generalization capability, yields visually satisfactory results and effectively removes most of the haze as depicted in Fig. 9j.

To assess the efficacy of UD³Net in UAV-based object detection, the YOLO11n model was employed to compare the visualization outcomes of object detection before and after dehazing with UD³Net. The number displayed above each bounding box represents the confidence level of the detection. As demonstrated in Fig. 10, utilizing hazed images for target detection can result in both false detections and missed detections. In contrast, the dehazed images produced by UD³Net not only substantially enhance the confidence levels but also significantly improve detection accuracy across various scenarios. In addition, Table 7 presents a quantitative comparison of detection performance before and after dehazing. The proposed dehazing framework significantly improves precision (P), recall (R), and mean Average Precision (mAP), achieving a 6.7% increase in mAP (56.9% vs. 63.6%).

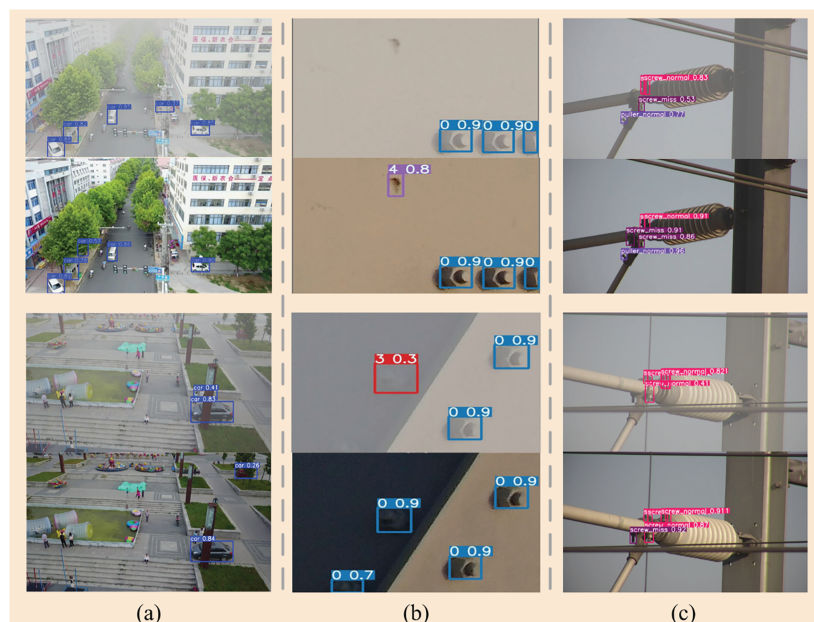


Figure 10: Performance comparison of traffic infrastructure target detection using YOLOv11 before and after dehazing. (a) Depicts the scenario of road vehicle detection. (b) Depicts the defect detection scenario for steel frame bridges, where label 0 denotes normal nuts, label 3 denotes rusted bolts, and label 4 denotes surface rust. (c) Depicts the scenario of high-speed rail contact network detection.

Table 7: Quantitative comparison of detection results before and after dehazing.

Class	Before (%)			After (%)		
	P	R	mAP	P	R	mAP
car	87.2	68.8	76.2	89.7	72.7	80.1
truck	55.6	25.3	28.2	60.5	33.5	37.7
bus	87	57.7	66.5	87.5	64.2	73.1
all	76.6	50.6	56.9	79.2	56.6	63.6

Note: P, R, and mAP represent Precision, Recall, and mean Average Precision, respectively.

In terms of model efficiency, UD³Net strikes a balance between effective haze removal and computational efficiency. The comparative results are presented in Table 8. UD³Net has fewer parameters and lower FLOPs than most models. Although some lightweight models, such as AODNet, demonstrate significant advantages regarding the number of parameters and FLOPs, they necessitate supervised training and exhibit inferior performance compared to our approach. In contrast to models that excel in outdoor real haze scenarios, such as PSD and RIDCP, UD³Net significantly reduces both the parameters and FLOPs.

Table 8: Model efficiency evaluating comparisons with paired and unpaired models.

	Method	Params (M)	FLOPs (Gmac)
Paired	AODNet	0.002	0.115
	FFANet	4.46	288.86
	PSD	33.11	182.50
	AECRNet	2.61	43.10
	C2PNet	7.17	463.49
Unpaired	CycleGAN	11.38	56.89
	RefineDNet	65.80	75.41
	YOLY	32.00	–
	RIDCP	29.48	193.43
	Ours	12.22	2.33

To evaluate the practical feasibility of the proposed method for real-time and high-resolution applications (such as UAV inspection), we analyzed the computational cost across different input resolutions. Table 9 details the FLOPs, Inference Time, and Frame Per Second (FPS) measured on a single NVIDIA RTX 3080ti GPU.

Table 9: Scalability analysis of the proposed method under different resolutions.

Resolution	FLOPs (G)	Inference Time (ms)	FPS
256 × 256	2.328	5.36	186.7
512 × 512	9.263	6.26	159.8
640 × 640	14.459	8.86	112.9
1280 × 720	33.216	18.60	53.7
1920 × 1080	73.535	39.02	25.6

The results demonstrate that the computational overhead of the proposed method remains manageable as resolution increases. At 1920×1080 resolution, the model still delivers 25.6 frames per second, validating its scalability for processing high-definition inputs. In terms of training efficiency, the model's total training duration on the ITS dataset was approximately 11.3 h, averaging 3 min per epoch. Therefore, we believe UD³Net can be applied to real-time UAV deployment.

5 Conclusion

This study presents an effective unsupervised dehazing framework, UD³Net. This framework is informed by frequency-domain prior and retinex theory during the training process and integrates contrastive learning to effectively constrain the features of the generated dehazed images. This approach mitigates the challenge of lacking ground truth in unsupervised learning and ensures that the model's output aligns with the corresponding clear images. To enhance the quality of the generated images, we designed a dual-domain convolutional mixer to facilitate the estimation of transmission maps and depth maps. In summary, UD³Net addresses the shortcomings of existing methods, specifically the absence of frequency-domain information and inadequate generalization capabilities. Experimental results demonstrate that UD³Net achieves commendable performance on both synthetic and real haze scene datasets.

Nonetheless, a significant limitation is observed in its prior guidance effectiveness under conditions of high haze density and during nighttime dehazing, which constrains the overall performance of UD³Net. This limitation primarily arises from the restricted information available in the images. Specifically, the Retinex-based illumination component (L) relies on the assumption of spatially smooth lighting variations. In night-time scenarios characterized by non-uniform, artificial point light sources, this assumption is often violated, causing L to misinterpret high-frequency glowing effects as global illumination and leading to local artifacts or under-enhanced regions. Additionally, since the phase spectrum primarily contains structural information, the assumption of "Phase Consistency" remains largely valid even in non-uniform fog. The network can leverage this invariant structural cue to recover details. However, the prior fails when the fog is extremely dense and acts as an occluder, where structural information is completely erased from the input. We anticipate that future research will further optimize its robustness and adaptability to improve performance in extreme environments.

Acknowledgement: Not applicable.

Funding Statement: This work was supported in part by The National Natural Science Foundation of China under Grant 52362048, in part by Yunnan Fundamental Research Projects under Grants 202301BE070001-042, 202401AT070409.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Shunpeng Yang, Cheng Yang and Yu Qian; methodology, Shunpeng Yang and Yunpeng Wu; software, Shunpeng Yang and Cheng Yang; formal analysis, Shunpeng Yang and Yunpeng Wu; investigation, Shunpeng Yang and Yunpeng Wu; resources, Yunpeng Wu; data curation, Shunpeng Yang and Yunpeng Wu; writing—original draft preparation, Shunpeng Yang and Yunpeng Wu; writing—review and editing, Shunpeng Yang, Yunpeng Wu, Wenwen Qin, Cheng Yang, and Yu Qian; visualization, Shunpeng Yang; supervision, Cheng Yang; project administration, Yunpeng Wu; funding acquisition, Yunpeng Wu. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the author [Shunpeng Yang, 20242206120@stu.kust.edu.cn].

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cao Z, Qin Y, Jia L, Xie Z, Liu Q, Ma X, et al. Haze removal of railway monitoring images using multi-scale residual network. *IEEE Trans Intell Transp Syst.* 2021;22(12):7460–73. doi:10.1109/TITS.2020.3003129.
2. Wu Y, Qin Y, Wang Z, Ma X, Cao Z. Densely pyramidal residual network for UAV-based railway images dehazing. *Neurocomputing.* 2020;371:124–36. doi:10.1016/j.neucom.2019.06.076.
3. Harish Babu G, Venkatram N. A survey on analysis and implementation of state-of-the-art haze removal techniques. *J Vis Commun Image Represent.* 2020;72(1):102912. doi:10.1016/j.jvcir.2020.102912.
4. He K, Sun J, Tang X. Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell.* 2011;33(12):2341–53. doi:10.1109/TPAMI.2010.168.
5. Zhu Q, Mai J, Shao L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans Image Process.* 2015;24(11):3522–33. doi:10.1109/TIP.2015.2446191.
6. Li CY, Guo JC, Cong RM, Pang YW, Wang B. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Trans Image Process.* 2016;25(12):5664–77. doi:10.1109/TIP.2016.2612882.
7. Ju M, Ding C, Ren W, Yang Y, Zhang D, Guo YJ. IDE: image dehazing and exposure using an enhanced atmospheric scattering model. *IEEE Trans Image Process.* 2021;30:2180–92. doi:10.1109/TIP.2021.3050643.
8. Ju M, Ding C, Guo CA, Ren W, Tao D. IDRLP: image dehazing using region line prior. *IEEE Trans Image Process.* 2021;30:9043–57. doi:10.1109/TIP.2021.3122088.
9. Berman D, Treibitz T, Avidan S. Non-local image dehazing. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016 Jun 27–30; Las Vegas, NV, USA. p. 1674–82. doi:10.1109/cvpr.2016.185.
10. Qin X, Wang Z, Bai Y, Xie X, Jia H. FFA-net: feature fusion attention network for single image dehazing. *Proc AAAI Conf Artif Intell.* 2020;34(7):11908–15. doi:10.1609/aaai.v34i07.6865.
11. Babu GH, Odugu VK, Venkatram N, Satish B, Revathi K, Rao BJ. Development and performance evaluation of enhanced image dehazing method using deep learning networks. *J Vis Commun Image Represent.* 2023;97:103976. doi:10.1016/j.jvcir.2023.103976.
12. Li C, Guo C, Guo J, Han P, Fu H, Cong R. PDR-net: perception-inspired single image dehazing network with refinement. *IEEE Trans Multimed.* 2020;22(3):704–16. doi:10.1109/TMM.2019.2933334.
13. Wu H, Qu Y, Lin S, Zhou J, Qiao R, Zhang Z, et al. Contrastive learning for compact single image dehazing. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2021 Jun 20–25; Nashville, TN, USA. p. 10546–55. doi:10.1109/cvpr46437.2021.01041.
14. Zheng Y, Zhan J, He S, Dong J, Du Y. Curricular contrastive regularization for physics-aware single image dehazing. In: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2023 Jun 17–24; Vancouver, BC, Canada. p. 5785–94. doi:10.1109/CVPR52729.2023.00560.
15. Wu X, Li Z, Guo X, Xiang S, Zhang Y. Multi-level perception fusion dehazing network. *PLoS One.* 2023;18(10):e0285137. doi:10.1371/journal.pone.0285137.
16. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*; 2017 Oct 22–29; Venice, Italy. p. 2242–51. doi:10.1109/ICCV.2017.244.
17. Yang Y, Wang C, Liu R, Zhang L, Guo X, Tao D. Self-augmented unpaired image dehazing via density and depth decomposition. In: *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2022 Jun 18–24; New Orleans, LA, USA. p. 2027–36. doi:10.1109/CVPR52688.2022.00208.
18. Wang Y, Yan X, Wang FL, Xie H, Yang W, Zhang XP, et al. UCL-dehaze: toward real-world image dehazing via unsupervised contrastive learning. *IEEE Trans Image Process.* 2024;33:1361–74. doi:10.1109/TIP.2024.3362153.
19. Yang Y, Wang C, Guo X, Tao D. Robust unpaired image dehazing via density and depth decomposition. *Int J Comput Vis.* 2024;132(5):1557–77. doi:10.1007/s11263-023-01940-5.

20. Zhao S, Zhang L, Shen Y, Zhou Y. RefineDNet: a weakly supervised refinement framework for single image dehazing. *IEEE Trans Image Process.* 2021;30:3391–404. doi:10.1109/TIP.2021.3060873.
21. Chen X, Fan Z, Li P, Dai L, Kong C, Zheng Z, et al. Unpaired deep image dehazing using contrastive disentanglement learning. In: *Computer vision—ECCV 2022*. Cham, Switzerland: Springer Nature; 2022. p. 632–48. doi:10.1007/978-3-031-19790-1_38.
22. Engin D, Genc A, Ekenel HK. Cycle-dehaze: enhanced CycleGAN for single image dehazing. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; 2018 Jun 18–22; Salt Lake City, UT, USA. doi:10.1109/CVPRW.2018.00127.
23. Lin K, Wang G, Li T, Wu Y, Li C, Yang Y, et al. Toward generalized and realistic unpaired image dehazing via region-aware physical constraints. *IEEE Trans Circuits Syst Video Technol.* 2025;35(3):2753–67. doi:10.1109/TCSVT.2024.3497594.
24. Qiao Y, Shao M, Wang L, Zuo W. Learning depth-density priors for Fourier-based unpaired image restoration. *IEEE Trans Circuits Syst Video Technol.* 2024;34(4):2604–18. doi:10.1109/TCSVT.2023.3305996.
25. Gao N, Jiang X, Zhang X, Deng Y. Efficient frequency-domain image deraining with contrastive regularization. In: *Computer vision—ECCV 2024*. Cham, Switzerland: Springer Nature; 2024. p. 240–57. doi:10.1007/978-3-031-72940-9_14.
26. Liu J, Wang S, Chen C, Hou Q. DFP-Net: an unsupervised dual-branch frequency-domain processing framework for single image dehazing. *Eng Appl Artif Intell.* 2024;136:109012. doi:10.1016/j.engappai.2024.109012.
27. Narasimhan SG, Nayar SK. Vision and the atmosphere. *Int J Comput Vis.* 2002;48(3):233–54. doi:10.1023/A:1016328200723.
28. Yu H, Zheng N, Zhou M, Huang J, Xiao Z, Zhao F. Frequency and spatial dual guidance for image dehazing. In: *Computer vision—ECCV 2022*. Cham, Switzerland: Springer Nature; 2022. p. 181–98. doi:10.1007/978-3-031-19800-7_11.
29. Xue M, Fan S, Palaiahnakote S, Zhou M. UR2P-Dehaze: learning a simple image dehaze enhancer via unpaired rich physical prior. *Pattern Recognit.* 2026;170:111997. doi:10.1016/j.patcog.2025.111997.
30. Ren W, Liu S, Zhang H, Pan J, Cao X, Yang MH. Single image dehazing via multi-scale convolutional neural networks. In: *Computer vision—ECCV 2016*. Cham, Switzerland: Springer International Publishing; 2016. p. 154–69. doi:10.1007/978-3-319-46475-6_10.
31. Cai B, Xu X, Jia K, Qing C, Tao D. DehazeNet: an end-to-end system for single image haze removal. *IEEE Trans Image Process.* 2016;25(11):5187–98. doi:10.1109/TIP.2016.2598681.
32. Li B, Peng X, Wang Z, Xu J, Feng D. AOD-net: all-in-one dehazing network. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*; 2017 Oct 22–29; Venice, Italy. p. 4780–8. doi:10.1109/ICCV.2017.511.
33. Guo C, Yan Q, Anwar S, Cong R, Ren W, Li C. Image dehazing transformer with transmission-aware 3D position embedding. In: *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2022 Jun 18–24; New Orleans, LA, USA. p. 5802–10. doi:10.1109/CVPR52688.2022.00572.
34. Song Y, He Z, Qian H, Du X. Vision transformers for single image dehazing. *IEEE Trans Image Process.* 2023;32:1927–41. doi:10.1109/TIP.2023.3256763.
35. Wang T, Zhang K, Shao Z, Luo W, Stenger B, Lu T, et al. GridFormer: residual dense transformer with grid structure for image restoration in adverse weather conditions. *Int J Comput Vis.* 2024;132(10):4541–63. doi:10.1007/s11263-024-02056-0.
36. Jiang X, Zhang X, Gao N, Deng Y. When fast Fourier transform meets transformer for image restoration. In: *Computer vision—ECCV 2024*. Cham, Switzerland: Springer Nature; 2024. p. 381–402. doi:10.1007/978-3-031-72995-9_22.
37. Sun J, Liu H, Wang Y, Zhang XP, Wei M. WDMamba: when wavelet degradation prior meets vision mamba for image dehazing. *arXiv:2505.04369*. 2025.
38. Wang Y, Chen L, Hu B, Liu H, Zhang XP, Wei M. Laplace-mamba: laplace frequency prior-guided mamba-CNN fusion network for image dehazing. *arXiv:2507.00501*. 2025.

39. Zhao Z, Qin Y, Qian Y, Wu Y, Qin W, Zhang H, et al. Automatic potential safety hazard evaluation system for environment around high-speed railroad using hybrid U-shape learning architecture. *IEEE Trans Intell Transp Syst.* 2025;26(1):1071–87. doi:10.1109/TITS.2024.3487592.
40. Wu Y, Zhao Z, Chen P, Guo F, Qin Y, Long S, et al. Hybrid learning architecture for high-speed railroad scene parsing and potential safety hazard evaluation of UAV images. *Measurement.* 2025;239:115504. doi:10.1016/j.measurement.2024.115504.
41. Li B, Gou Y, Gu S, Liu JZ, Zhou JT, Peng X. You only look yourself: unsupervised and untrained single image dehazing neural network. *Int J Comput Vis.* 2021;129(5):1754–67. doi:10.1007/s11263-021-01431-5.
42. Chen Z, Wang Y, Yang Y, Liu D. PSD: principled synthetic-to-real dehazing guided by physical priors. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA.* p. 7176–85. doi:10.1109/cvpr46437.2021.00710.
43. Wu RQ, Duan ZP, Guo CL, Chai Z, Li C. RIDCP: revitalizing real image dehazing via high-quality codebook priors. In: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada.* p. 22282–91. doi:10.1109/CVPR52729.2023.02134.
44. Narasimhan SG, Nayar SK. Chromatic framework for vision in bad weather. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition CVPR 2000; 2000 Jun 15; Hilton Head, SC, USA.* p. 598–605. doi:10.1109/CVPR.2000.855874.
45. Dong H, Pan J, Xiang L, Hu Z, Zhang X, Wang F, et al. Multi-scale boosted dehazing network with dense feature fusion. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA.* p. 2154–64. doi:10.1109/cvpr42600.2020.00223.
46. Schar H. *Optimal operators in digital image processing [dissertation].* Heidelberg, Germany: Heidelberg University; 2000.
47. Ye T, Jiang M, Zhang Y, Chen L, Chen E, Chen P, et al. Perceiving and modeling density is all you need for image dehazing. arXiv:2111.09733. 2021.
48. Chen Z, He Z, Lu ZM. DEA-net: single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Trans Image Process.* 2024;33:1002–15. doi:10.1109/TIP.2024.3354108.
49. Lu L, Xiong Q, Xu B, Chu D. MixDehazeNet: mix structure block for image dehazing network. In: *Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN); 2024 Jun 30–Jul 5; Yokohama, Japan.* p. 1–10. doi:10.1109/IJCNN60899.2024.10651326.
50. Land EH. The retinex theory of color vision. *Sci Am.* 1977;237(6):108–29. doi:10.1038/scientificamerican1277-108.
51. Gui J, Cong X, He L, Tang YY, Kwok JT. Illumination controllable dehazing network based on unsupervised retinex embedding. *IEEE Trans Multimed.* 2024;26:4819–30. doi:10.1109/TMM.2023.3326881.
52. Cui Y, Wang Q, Li C, Ren W, Knoll A. EENet: an effective and efficient network for single image dehazing. *Pattern Recognit.* 2025;158:111074. doi:10.1016/j.patcog.2024.111074.
53. Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP. Least squares generative adversarial networks. In: *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy.* p. 2813–21. doi:10.1109/ICCV.2017.304.
54. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. 2014.
55. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, et al. Benchmarking single image dehazing and beyond. *IEEE Trans Image Process.* 2018;28(1):492–505. doi:10.1109/TIP.2018.2867951.
56. Feng C, Chen Z, Li X, Wang C, Yang J, Cheng MM, et al. HazyDet: open-source benchmark for drone-view object detection with depth-cues in hazy scenes. arXiv:2409.19833. 2024.
57. Choi LK, You J, Bovik AC. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Trans Image Process.* 2015;24(11):3888–901. doi:10.1109/TIP.2015.2456502.
58. Mittal A, Moorthy AK, Bovik AC. Blind/referenceless image spatial quality evaluator. In: *Proceedings of the 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR); 2011 Nov 6–9; Pacific Grove, CA, USA.* p. 723–7. doi:10.1109/ACSSC.2011.6190099.
59. Mittal A, Soundararajan R, Bovik AC. Making a “completely blind” image quality analyzer. *IEEE Signal Process Lett.* 2013;20(3):209–12. doi:10.1109/LSP.2012.2227726.